The variable influence of confession inconsistencies: How factual errors (but not contradictions) reduce belief in suspect guilt

Journal:	*Applied Cognitive Psychology*
Manuscript ID	ACP-20-0019.R1
Wiley - Manuscript type:	Research Article
Keywords:	confessions, wrongful conviction, juror decisions, attribution, suspicion

RUNNING HEAD: CONFESSION INCONSISTENCIES

The variable influence of confession inconsistencies: How factual errors (but not contradictions) reduce belief in suspect guilt

Word count: 6,629 (excluding abstact, tables, figures, and references)

Abstract

Wrongful conviction statistics suggest that jurors pay little heed to the quality of confession evidence when making verdict decisions. However, recent research indicates that confession inconsistencies may sometimes reduce perception of suspect guilt. Drawing on theoretical frameworks of attribution theory, correspondence bias, and the story model of juror decision-making, we investigated how judgments about likely guilt are affected by different types of inconsistencies: self-contradictions (Experiment 1) and factual errors (Experiment 2). Crucially, judgments of likely guilt of the suspect were reduced by factual errors in confession evidence, but not by contradictions. Mediation analyses suggest that this effect of factual errors on judgments of guilt is underpinned by the extent to which mock-jurors generated a plausible, alternative explanation for why the suspect confessed. These results indicate that not all confession inconsistencies are treated equally; factual errors might cause suspicion about the veracity of the confession, but contradictions do not.

Keywords: confessions, wrongful conviction, juror decisions, attribution, suspicion

In 1979, six-year-old Etan Patz left home to catch the school bus alone for the first time and was never seen again. Although presumed kidnapped and murdered, no physical evidence of the crime was ever recovered, with the prosecution case relying on sometimes contradictory witness and confession statements. The man convicted of Etan's disappearance and murder, Pedro Hernandez, was 18 years old at the time of Etan's disappearance and worked at the bodega near the bus stop where Etan was last seen. Hernandez confessed to family and friends in the 1980s that he had done something terrible, and then made further confessions to police in 2012 about his involvement in Etan Patz's disappearance, but repeatedly denied murdering him. Hernandez later retracted those confessions, citing issues of mental illness, and pleaded not guilty. After 18 days of jury deliberation, the belief in Hernandez's guilt moved from 50:50 guilty/not guilty, to a final result of 11 jurors rendering a guilty verdict, with just one holdout juror unwilling to move from a not guilty verdict. That holdout juror stated that he could not move beyond reasonable doubt due to the bizarre and inconsistent confessions that the case relied heavily upon, though stopped short of stating that he believed Hernandez to be innocent. Conversely, another juror stated that it was the quality and quantity of Hernandez's confessions that eventually swayed his verdict to guilty. Whether Hernandez's confessions were true or false, the differences in juror perception of confession quality is crucial to understanding why some jurors are willing to convict based on inconsistent confession evidence, while others remain skeptical of the confession's usefulness as a piece of evidence when its veracity is in question. As well as real life cases where inconsistent confession evidence is questioned by individual jurors, emerging research has also shown that some jurors are able to evaluate confession evidence more objectively than previously thought (e.g., Palmer, Button, Barnett, & Brewer, 2016; Woestehoff & Meissner, 2016). The current research furthers this research by investigating the conditions under

which inconsistencies in confession evidence prompt people to find confessions less compelling when making verdict decisions.

Although an inconsistent confession is no guarantee that the person is innocent, false confessions are surprisingly common, despite their counter-intuitive nature. Statistics show that people falsely confess at rates more frequent than common sense might indicate, with approximately one quarter of wrongful convictions overturned by DNA evidence attributed at least partly to an innocent person falsely confessing their involvement in the crime ("Innocence Project" 2017). These numbers highlight two important issues: why people falsely confess, and why juries are susceptible to accepting those false confessions. As to why people falsely confess, a large body of research has identified interrogation tactics that pressure confessions from guilty and innocent suspects alike (e.g. Leo, 2008; Kassin, 2008), such as lengthy interrogations (Leo, 2008), and using false evidence (Kassin, 2008). This research has led to important recommendations for policies to reduce the likelihood of false confessions occurring (e.g. Kassin et al., 2010), though the uptake of the recommendations is sometimes slow. For example, by 2018, only 25 US states had implemented the requirement for video recording of custodial interrogations (Bang, Stanton, Hemmens, & Stohr, 2018). Resistance to procedural change, and the widespread use of pressurised interrogation techniques that increase the likelihood of false confessions—such as the Reid technique (Kassin, 2008)—suggest that false confessions will continue to occur, and be presented in court. Hence, it is crucial to understand how jurors process confession evidence in order to determine the conditions under which false confessions are likely to translate to guilty verdicts.

The power of confession evidence

Confessions make for compelling evidence, being described as the most powerful piece of courtroom testimony that a prosecutor could present to secure a conviction (Kassin

& Neumann, 1997). The strength of that confession is such that it can hold almost the entire prosecution case, with little corroborating evidence needed (Leo, 2013), and can result in otherwise contradictory evidence being glossed over or ignored completely by jurors (Malloy & Lamb, 2010). The persuasive capacity of confessions, and why they hold such power in court, relates to beliefs about how and why people confess. Logically, people believe that a person confesses because they are guilty (Henkel et al., 2008; Kassin et al., 2010). Given the negative consequences of confessing to a crime, it can be difficult to imagine a plausible reason as to why a person would confess if not guilty. Leo and Ofshe (1998) describe this phenomenon as the *myth of psychological interrogation*: the belief that a person of sound mind would not confess to a crime they did not commit, unless physically tortured. This implies that jurors cannot imagine a reason (other than guilt) as to why a suspect would confess. Such jurors would then be susceptible to accepting a confession at face value, and attributing variations in confession quality to reasons other than a lack of complicity in the crime. This notion is supported by the results of studies showing that jurors are willing to convict based on a confession even when the confession was coerced (Kassin & Wrightsman, 1981), when the confession was provided by an informant with incentive to lie (Neuschatz, Lawson, Swanner, Meissner, & Neuschatz, 2008), or in total disregard of exculpatory DNA evidence (Appleby & Kassin, 2016).

One reason why jurors might fail to reject false confession evidence relates to the apparent low level of scrutiny afforded to confessions, in contrast to other types of evidence. There is some evidence that jurors tend to take confessions at face value, and do not treat confession evidence with the same level of scepticism and scrutiny as they would when assessing the quality and motives of witness or victim testimony (Malloy & Lamb, 2010; Palmer et al., 2016; Redlich, Ghetti, & Quas, 2008; Woestehoff & Meissner, 2016). In an overview of the way in which jurors process different evidence types, Malloy and Lamb

(2010) observed that jurors often ignore the same type of inconsistencies in confession evidence that would have caused them to believe the testimony lacking in credibility and reliability had it come from an eyewitness (e.g., Berman, Narby, & Cutler, 1995). This could be because inconsistencies are often interpreted as deceit, such as in the case of a witness who changes their story over time (Malloy & Lamb, 2010), and can reduce the credibility of that witness. However, inconsistencies in confessions present a more complicated picture as there is no obvious benefit for an innocent person to confess, and may therefore act to reduce perceptions of guilt.

However, recent research (Alceste, Crozier, & Strange, 2019; Henderson & Levett, 2016; Eric Jones, Bandy, & Palmer Jr, 2019; Palmer et al., 2016; Woestehoff & Meissner, 2016) suggests that jurors pay more attention to confession inconsistencies than wrongful conviction statistics might suggest. These studies show that some inconsistencies in confession evidence do indeed reduce perceptions of guilt of the confessor. For example, Palmer et al. (2016) had participants in a mock-juror study read materials that included a confession statement of a suspect admitting to armed robbery. The confession was manipulated to be factually consistent or inconsistent with the facts of the crime, and participants were either provided with an alternate reason why the suspect might have confessed (police coercion, or to protect someone), or no alternate reason was provided. Participants who read a confession statement inconsistent with the facts of the crime were less likely to give a guilty verdict than those who read a confession statement that factually matched information about the crime. This result was found whether a plausible alternate explanation for the suspect's confession was made salient to the juror or not. These findings indicate that some circumstances may motivate jurors to generate their own reason why the suspect has confessed, despite the general belief that jurors fail to scrutinise confession veracity.

Alceste et al., 2019 similarly tested the effect of inconsistencies and alternate reasons for confessions on perception of suspect guilt. They found that inconsistent confessions, and those where key crime information was provided by the interrogator rather than the suspect, influenced belief in the presence of an alternate reason for the confession (such as coercion). The findings of Alceste et al., 2019 and Palmer et al., 2016 show that participants are willing to attribute confessions to a reason other than guilt. However, these studies do not explicitly test if this effect is still present when no salient reason is provided to participants. The present study aims to investigate the capacity of jurors to generate their own alternate explanations for an inconsistent confession, and how this might influence belief in suspect guilt. Addressing this gap in the literature will help further understanding of some of the conditions that determine when—and how—different confession inconsistencies affect belief in suspect guilt.

Theoretical rationale

Previous research suggests that the likelihood of jurors generating alternative explanations for why a suspect confessed (other than because the suspect was guilty) may play an important role in explaining why inconsistencies in confessions reduce jurors' perceptions of guilt in some situations but not others (Alceste et al., 2019; Palmer et al., 2016; Woestehoff & Meissner, 2016). One reason why confessions are extremely persuasive of guilt relates to the *correspondence bias* (e.g. Gilbert & Malone, 1995; Ross, 1977). Jurors tend to attribute confessions to internal factors (i.e., guilt) rather than external factors, such as situational pressures associated with police interrogations (e.g., Kassin, 2012; Woestehoff & Meissner, 2016). The correspondence bias is especially likely to affect jurors' processing of confessions because it is exacerbated for behaviors that are not self-serving, such as confessing to a crime (e.g. Edward Jones & Davis, 1965; Kassin, 2012; Kelley, 1973).

Crucially, observers are less likely to attribute behaviors to internal factors when there is a salient alternative explanation for the behavior (Fein, McCloskey, & Tomlinson, 1997; Fein, Morgan, Norton, & Sommers, 2010; Kelley, 1973). Thus, if jurors are able to think of a plausible, alternative reason why a person has confessed (other than being guilty of the crime), they will be less likely to attribute the confession to guilt rather than some other cause (e.g., the suspect was protecting the real perpetrator, or was pressured by police to confess). The story model of juror decision making (Pennington & Hastie, 1992) is particularly relevant in this context, as it proposes that jurors create a cohesive narrative as the trial unfolds, and that this narrative is necessarily adaptable to accommodate new or conflicting evidence. The story model may then provide some foundation for the idea that, when faced with an inconsistent confession, some jurors generate an alternate explanation as to why the suspect has confessed in order to fill a gap in the narrative created by the loss of the confession as a valid piece of evidence.

Based on this reasoning, we investigated whether the effect of inconsistencies on peoples' belief in guilt depends on the extent to which inconsistencies prompt people to generate a plausible alternative explanation for why the suspect confessed. We hypothesized that the generation of alternative explanations for the confession would mediate the effect of inconsistencies on belief in guilt. That is, that the presence of inconsistencies would influence the generation of plausible alternative explanations for the confession (i.e., reasons other than guilt as to why the person confessed) which, in turn, would reduce belief in guilt ratings.

In the context of this broad aim, we addressed two additional issues. First, we tested whether different types of confession inconsistencies (*contradictions* vs *factual errors*) differed in the extent to which they prompted people to generate alternate explanations and, hence, the extent to which they influenced belief in guilt. Factual errors involve the suspect making statements that are independently refuted by a piece of verified evidence, such as a

police report of the facts of the crime, or photographs of the crime scene. Contradictions occur when the suspect seems to forget the facts of their own story and must correct themselves in order to keep the story coherent. The latter are in line with Brewer and Hupfeld's (2004) definition of the inconsistency being due to distraction rather than error.

In two decision-making experiments, we compared consistent confession evidence with evidence containing either contradictions (Experiment 1), or factual errors (Experiment 2). Transcripts for the two experiments contained confession inconsistencies that focused around key aspects of the crime, including time and place, rather than peripheral aspects which jurors might see as inconsequential when determining the suspect's guilt. If some types of inconsistencies are more likely to facilitate generation of alternate explanations, this could contribute to the variable effects of inconsistencies on the persuasiveness of confession evidence reported in the literature. For example, contradictions might be viewed as natural irregularities in the confession story that simply reflect harmless imperfections in the confessor's memory or recounting of the crime. If this is the case, contradictions might not trigger any concerns about the confession or prompt the jurors to consider any alternative explanation—other than guilt—for why the confession was made. In contrast, factual errors could be perceived as indicating serious problems with the credibility of a confession (e.g., *"If the person was really there, surely they would not have got that detail wrong?"*). Hence, factual errors might prompt jurors to consider alternative explanations for the confession which, in turn, could reduce perceptions of guilt. To the extent that this reasoning holds true, factual errors—but not contradictions—would increase the likelihood of jurors generating a plausible alternate explanation for the confession.

Recent research by Jones et al. (2019) is consistent with part of our contemplation of how jurors might process confession evidence. Jones et al. found that factual errors in a confession led to mock-jurors giving reduced judgments of the probable guilt of the

confessor. Although the authors did not measure the generation of alternative explanations for why the defendant confessed, they did find that factual errors increased perceptions that the confession was coerced, and of the likely guilt of an accomplice. Jones et al. suggested that these findings might indicate that mock-jurors thought the defendant was confessing to cover for the accomplice; this would amount to generating an alternative explanation for the confession. While speculative, Jones et al.'s interpretation of results is consistent with our own hypotheses about the role of juror-generated alternate explanations in response to inconsistent confession information. We tested the hypothesis explicitly by comparing consistent confession evidence with evidence that contained contradictions (Experiment 1) or factual errors (Experiment 2) and measuring the extent to which each type of inconsistency prompted the generation of alternative explanations for the confession.

Second, we considered the specific mechanism by which the generation of alternative explanations might translate to differences in guilt ratings. One possibility is that this effect hinges on the number of alternative explanations generated. That is, the greater the number of reasons a juror can think of (other than guilt) for why the suspect confessed, the lower the perception of guilt. Another possibility is that the effect is reliant on the subjective plausibility of the best alternative explanation, rather than the number of explanations generated. That is, regardless of how many alternative explanations a juror generates, belief in guilt will depend on the extent to which at least one highly plausible explanation is generated—the greater the plausibility of the best explanation, the lower the perception of guilt.

The investigation of the role of plausibility of self-generated alternate explanations for a confession has, to our knowledge, not been studied previously, and is our key contribution to the literature. Previous literature has addressed different variables that might act as alternative explanations for a confession, such as coercion (e.g. Jones et al., 2019; Palmer et

al., 2016; Henderson & Levett, 2016), and whether the interrogator or the suspect provided confession details (Alceste et al., 2019). However, these alternate explanations were provided to participants, and therefore did not explicitly test whether people are able to spontaneously generate their own explanations for a confession, and the subsequent role of plausibility of belief in those alternate explanations.

We designed a task to measure (1) the number of alternative explanations jurors could generate for why the suspect confessed, and (2) the plausibility of the best alternative explanation they generated. After providing a verdict (guilty or not guilty), participants were asked to list any other reasons why the suspect may have confessed (other than guilt). If the participant generated at least one alternate explanation for the confession, they were asked to choose the most plausible of their explanations and rate its plausibility.

Experiment 1: Contradictions

Experiment 1 investigated whether inconsistencies in the form of contradictions (e.g., Brewer & Hupfeld, 2004) would prompt the generation of alternate explanations for why a suspect confessed that would then, in turn, influence belief in guilt. If contradictions in confession evidence increase the likelihood of generating a plausible alternate explanation, belief in guilt should be reduced.

Method

Materials and data from Experiments 1 and 2 are available from the authors on request.

Participants

Seventy-three participants (53 female) were recruited from undergraduate psychology courses in return for partial course credit. Participant ages ranged from 18 to 60 years (M = 27.47, SD = 11.17), and were either reimbursed AUD$15, or received partial course credit.

Design and procedure

Informed consent was obtained prior to random allocation of participants to one of two conditions (*consistent, contradictions*) of a between-groups design. Supervised testing took place in groups (n = 2 - 12) in a laboratory or classroom setting. After giving informed consent, participants read a confession interview transcript and then completed a pen-and-paper questionnaire, with instructions not to change their answers once given. In order to understand the effect of confession inconsistencies on individual belief in guilt, rather than as an indication of how jurors would actually vote post-deliberation, participants acted as individuals, rather than as part of a juror deliberation exercise. Testing took approximately 25 minutes.

Transcript

The stimulus material was a (fictional) transcript of a police interview in which a suspect confessed to a physical assault charge. A fact-finding police interview (as per the PEACE model) was used in preference to a guilt-presumptive interrogation (such as the Reid technique) to avoid introducing coercion as an experimental artefact. At the start of the interview, it is stated that the suspect has previously spoken to the police, but that a new officer has taken over the case and wants to hear the suspect's story from the start: *"I'm just going to put your original statement aside for the moment, and give you the chance to start again from the beginning"*. In the *consistent* condition the suspect does not make any mis-steps in his story, and the police officer does not query the consistency of his account of the crime. In contrast, in the *contradictions* condition consistency was manipulated by having the suspect contradict the details from his previous statement on five key facts: 1) time the crime took place, 2) point of entry, 3) movements within the house, 4) injuries to the victim, and 5) whether he moved the victim or not. The contradictions between the suspect's current statement and the alluded-to previous statement, are pointed out by the police officer, forcing

the suspect to backtrack and amend his story to fit with his previously given statement. For example, the suspect stated that he entered the victim's house through an unlocked garage door. When the police officer asked if that agrees with his previous statement, the suspect responded, *"No, that's right. It was the other way round. I got into the house by the front door, but I left through the garage."* The suspect offers no clear reason as to why he makes these mistakes.

Measures

Participants were asked a number of case-related questions, including whether they believed the suspect guilty or not guilty of the assault charge (verdict), and how confident they were in their decision, from 1 ("*not at all confident"*) to 10 ("*totally confident"*). Verdict and verdict confidence were combined to form a new dependent variable, called 'belief in guilt', following the method outlined by Tenney, MacCoun, Spellman, and Hastie (2007). A value of 0.5 was added to each confidence score, and the scores for "not guilty" verdicts were multiplied by -1 resulting in a range of scores from -9.5 (completely confident in a not guilty verdict) to 9.5 (completely confident in a guilty verdict), with the 0.5 difference reflecting that the original confidence scale ran from 1 to 10, rather than 0 to 10. For example, a guilty verdict with a confidence rating of 4 became a belief in guilt of 3.5, while a not guilty verdict with a confidence rating of 8 became a belief in guilt of -7.5.

Manipulation checks of consistency and coercion asked participants to rate on a 10-point scale how consistent they found the suspect's evidence (very inconsistent, to totally consistent), and how voluntary they believed the suspect's confession to be (totally involuntary, to totally voluntary). Demographic information included age, gender, languages spoken at home, and whether the participant was studying at university full or part-time.

Alternate explanations and plausibility ratings.

The questionnaire included a measure of participants' ability to generate plausible alternate explanations for why the defendant confessed. This took part in two stages. First, participants were asked to imagine that the suspect was innocent of the confessed crime and list any reasons why the suspect might have confessed. For example, participant-generated explanations included, *"If he did something worse and was using this as an alibi"*, and *"Could be covering up for a mate or brother"*. Second, participants were asked to rate their preferred alternate explanation (if they gave one) as to how convincing they thought that reason was in explaining why the suspect confessed (on a scale of 0% - *this reason is not at all convincing*, to 100% - *this reason is totally convincing*). The score from the favored explanation was used as a measure of the extent to which participants were able to generate a plausible alternate explanation (other than guilt) for why the suspect confessed. Participants who did not provide any alternate explanations were coded with a score of zero for this variable.

Results

Manipulation checks

The consistency manipulation check confirmed that participants were able to discern the inconsistencies in the confession transcript. Participants rated a *consistent* confession as significantly more consistent than a contradictory confession, t (66.93) = 9.71, p <.001, 95% CI [3.37, 5.11], d = 2.27. A manipulation check of perceived confession voluntariness showed little difference in perceived voluntariness between the *consistent* and *contradictions* conditions, $t < 1$, d =.07 (see Table 1 for means, standard deviations and 95% confidence intervals around the means for Experiment 1 measures).

Generation of alternate explanations

There was no significant difference in the number of alternate explanations generated between the *consistent* and *contradictions* conditions, $t < 1$, $d = .17$. Participants who read a consistent confession gave similar plausibility ratings for their favored alternate explanation to those who read a contradictory confession, $t < 1$, $d = .12$.

Belief in guilt

An independent samples t-test revealed no effect of consistency on belief in guilt, $t < 1$, $d = 0.08$. Participants in both conditions overwhelmingly judged the suspect to be guilty of the confessed crime.

Dichotomous verdicts of guilt

Overall, 90.4% of participants believed the suspect was guilty. There was no significant difference in guilty verdicts between those in the *consistent* (93.5% guilty) and the *contradictions* conditions (88.1% guilty), $n = 73$, $\chi^2(1) = .612$, $p = .43$.

[Table 1 near here]

Discussion

Experiment 1 results indicate that while jurors were aware of contradictions in a confession, these contradictions were not sufficient to reduce belief in guilt. More specifically, jurors may perceive that the suspect is presenting a contradictory confession for reasons other than innocence. For example, if the suspect changes their story, or stumbles over key facts, this may be interpreted as an understandable imperfection in the retelling of a story, rather than a problem suggestive of the suspect's innocence and subsequent fabrication of a confession to escape the pressure of interrogation.

Experiment 2: Factual Errors

Experiment 2 tested whether factual errors in a confession (e.g., Jones et al., 2019) would prompt participants to generate plausible alternate explanations for the confession and, in turn, reduce their belief in the suspect's guilt.

Method

Participants and design

Eighty-nine participants (60 female) were recruited from undergraduate courses, and the wider university community. Participants were aged between 18 and 61 years (M = 27.19, SD = 10.44), and were either reimbursed AUD$15, or received partial course credit.

Design and procedure

Design and procedure are as per Experiment 1, with the exception that participants in Experiment 2 were allocated to either the *factual errors*, or *consistent* conditions of a between-groups design.

Materials

The confession transcript closely follows that of Experiment 1, though the reference to the suspect giving a statement previously was removed as this is not appropriate for Experiment 2.

In this experiment, it was necessary to include a police report alongside both the *consistent* and *inconsistent (factual errors)* conditions to allow participants to check for accuracy themselves. Unfortunately this meant that it was not possible to run Experiment 1 and 2 concurrently by creating a three-groups design (i.e. *consistent, contradictions, factual errors)*, as it was important that participants in Experiment 1 make their decisions about confession veracity based only on the information given by the suspect (whether consistently, or with contradictions), with no ability to fact check via a police statement. Had we included

the police statement in all conditions (i.e. including the contradictions condition), this could have created the perception of factual errors between the police statement and the confession, rather than restricting the participant to the suspect's intra-confession inconsistencies. Therefore, to avoid confounding our two different manipulations, Experiments 1 and 2 were run separately.

Experiment 2 participants in the consistent condition read a confession transcript that matched the police report on all key facts. In the *factual errors* condition, the transcript deviated from the police report on five key facts: 1) the time of the assault, 2) location of injury, 2) location where the assault took place, 3) point of entry to the house, and 4) location of the victim. Given that Alceste et al., (2019) found that the source of the confession detail could alter how participants perceived the veracity of confession, it is important to note that in Experiment 2 the suspect provides all of the key confession details (and subsequent errors), rather than the interviewing officers.

Measures

Measures were the same as for Experiment 1.

Results and discussion

Manipulation checks

Confession consistency was successfully manipulated, with participants rating a *consistent* confession rating as more consistent than a confession containing *factual errors*, t (75.343) = 10.00, $p < .001$, $d = 2.11$. There was no significant difference between conditions on the manipulation of voluntariness, t (87) = 1.30, $p = .197$, indicating that the manipulation did not inadvertently affect perceived voluntariness of the confession (see Table 2 for means, standard deviations and 95% confidence intervals around the means for Experiment 2 measures).

Dichotomous verdicts of guilt

Overall, 69.7% of participants believed the suspect to be guilty. Participants in the *factual errors* condition gave a significantly lower amount of guilty verdicts (51.1%), than those in the *consistent* condition (88.6%), $n = 89$, $\chi^2 (1) = 14.823$, $p < .001$. This result replicated the central result of Jones et al. (2019).

Belief in guilt

Consistent with Jones et al. (2019), an independent samples t-test revealed that participants who read a confession containing factual errors had a weaker belief in suspect guilt than did those who read a confession that was factually consistent with the police report, $t(77.21) = 4.15$, $p < .001$, $d = 0.88$. Note that this effect of inconsistency of belief in suspect guilt was not found in Experiment 1 where confession inconsistencies took the form of contradictions.

Generation of alternate explanations

The presence of factual errors affected belief in the plausibility of the favored alternate explanation, but not the number of alternate explanations generated. Participants in the *factual errors* condition expressed significantly higher belief in the plausibility of their favored alternate explanation than those in the *consistent* condition, $t (87) = -2.27$, $p = .025$, $d = .48$. However, there was no significant difference between conditions on the number of alternate explanations generated, $t (87) = -.80$, $p = .427$, $d = .17$.

We conducted a mediation analysis in order to investigate the mechanism by which confession inconsistency (in the form of factual errors) affected belief in guilt. Mediation analysis, using PROCESS software (Hayes, 2013), confirmed that the effect of inconsistencies on belief in guilt was partially mediated by the plausibility of the favored alternate explanation for the suspect's confession (see Fig. 2). The presence of factual errors increased ratings of the favored alternate explanation which, in turn, was associated with

lower belief in guilt, B = -.047, p=.021, *CI 95%* [-.09, -.01]. Most importantly, there was an indirect effect of inconsistency on belief in guilt via differences in plausibility (B = -.72, [-2.22, -.064]), showing that consistency affected belief in guilt by altering belief in the plausibility of the participant's favoured alternate explanation for the confession. The direct effect of inconsistency on belief in guilt (controlling for plausibility of favored alternate explanations) was statistically significant, indicating partial mediation.

[Figure 1 near here]

Crucially, while the plausibility rating of favored alternate explanations had a significant mediating effect on belief in guilt, the number of alternate explanations generated did not, B = -.17, [-1.28, .12]. This indicates that a participant's belief in the suspect's guilt is not affected by the number of alternate explanations they can think of, but rather how strongly they believe that their favoured alternate explanation (e.g. that the suspect was covering for someone else) can adequately explain why the suspect might have confessed in the absence of guilt.

[Table 2 near here]

General Discussion

In this research, we used psychological theory to examine whether individuals could discern the presence of inconsistencies in confession evidence, and if these inconsistencies then influenced their belief in the suspect's guilt. This is an important issue because it can help us understand factors that shape the extent to which jurors are sceptical when evaluating confession evidence. The findings of these two experiments make several contributions to advancing our understanding of how and when confession inconsistencies influence belief in suspect guilt.

First, our results highlight a crucial mechanism by which errors in a confession reduce belief in guilt: The extent to which the person judging the suspect's guilt can generate a

plausible explanation—other than guilt—for why the confession was made. When inconsistencies in confession evidence prompted participants to generate alternate explanations for why the suspect confessed (e.g., that the suspect confessed to cover for the person who actually committed the crime), this translated to reduced perceptions of guilt. This finding is encouraging because it contradicts the correspondence bias that says that people are incapable of attributing confessions to anything other than internal reasons (i.e. guilt), and cannot imagine why an innocent person would confess to a crime they did not commit. However, in considering how correspondence bias might play a role in jurors' processing of errors in confession evidence, one problem that arises is determining what specific target behavior the observers (i.e. jurors) are making attributions about. In confession research, it is usually jurors making attributions about the confession itself (e.g. Woestehoff & Meissner, 2016). However, when the confession contains errors, it is unclear whether the target behavior for making attributions is the confession or the errors themselves. There is no simple way of determining which is the target behavior, because the kinds of attributions made about the two behaviors (i.e. the confession, or the errors) are quite different. For example, if the confession is the target behavior and jurors are considering 'why did this person confess to a crime?', dispositional attributions are very likely to be negative (e.g., 'because they are evil, and they did it'). In contrast, if errors in the confession are the target behavior, jurors would be considering 'why did this person make mistakes in their confession?'. In this case, dispositional attributions will likely be much less negative, e.g., 'their memory is poor', or 'maybe they lied about committing the crime' (assuming that lying is viewed as less bad than committing a crime).

While our study does not allow us to differentiate whether participant attributions are related specifically to either the confession or the errors in that confession, our findings do show that people are indeed capable of generating alternate explanations for why the suspect

might have confessed (e.g., that the suspect was covering for someone else), and then using those explanations to alter their internal narrative of the crime and reduce their belief in the suspect's guilt. Second, our results show that the effect of inconsistencies on judgments if guilt varies depending on the nature of the inconsistencies. Replicating the results of Jones et al. (2019), factual errors in confession evidence were associated with lower ratings of guilt. In contrast, contradictions were not associated with lower guilt ratings. Importantly, these different types of inconsistencies differed not only in their effects on perceived guilt, but also in their effects on the generation of alternative explanations for the confession. The presence of factual errors increased the likelihood that participants would generate alternative explanations which, in turn, reduced judgments of guilt. In contrast, the presence of contradictions did not affect the generation of alternative explanations. Together, these results suggest that factual errors were treated by participants as indicators of a potentially problematic confession, whereas contradictions in confession evidence seem to have been disregarded as inconsequential. This difference might contribute to explaining why inconsistencies in confessions are sometimes—but not always—disregarded by jurors (e.g., Malloy & Lamb, 2010; Palmer et al., 2016). However, it still cannot completely explain this issue, because there are numerous cases of false confessions containing clear factual errors.

Third, our data show that the number of alternate explanations generated does not matter. Regardless of how many alternate explanations were generated, the extent to which inconsistencies translate to reduced belief in guilt depended on the generation of a single, plausible explanation. The more plausible this best alternative explanation was, the greater the reduction in perceived guilt of the defendant. This finding that competing alternative explanations did not have additive effects on belief in guilt may have implications for attribution-based theories of social judgment (e.g., Kelley, 1973; Fein et al., 1997). It is important to note that the effect of factual errors on belief in guilt was only partially mediated

by differences in the plausibility of favored alternate explanations for why the suspect confessed. Future research might identify additional mechanisms that contribute to the relationship between confession inconsistencies and belief in guilt. Nevertheless, our results indicate that the generation of a single, plausible alternate explanation plays an important role in explaining this relationship.

Finally, our findings exclude two possible explanations for why jurors are poor at processing confession evidence. The first is that jurors simply do not notice confession inconsistencies in the way that they would if the testimony was from an eyewitness. Similar to previous research (e.g. Alceste et al., 2019), participants in both experiments were consistently able to discern the inconsistencies present in confession evidence. The second explanation ruled out is that people are incapable of imagining why an innocent person would confess to a crime they did not commit. Contrary to this view, our results showed that mock jurors are capable of generating alternate explanations for why the suspect might have confessed (e.g., that the suspect was covering for someone else), and then use those explanations to alter their internal narrative of the crime and reduce their belief in the suspect's guilt (depending on the extent to which they generated a plausible alternate reason).

Limitations and future directions

We acknowledge that the evidence presented to mock-jurors in this research was much less complex than the entire body of evidence the jurors consider in many real cases. However, the aim of the methodology in present research was not to simulate the complexities of a real trial. Nor do we claim to account for all possible mechanisms that determine whether a juror will accept or reject an inconsistent confession. Rather, our purpose was to isolate one possible variable that might account for some of the variation in juror decisions about confession evidence (for detailed discussion of different methods used in juror decision research, see Bornstein, 1999; Kerr & Bray, 2005; Palmer, Horry, & Brewer,

2011). By removing the confession evidence from the noise of other trial evidence, we can test whether confession evidence alone holds the kind of courtroom power indicated by real cases. Much as the majority of cases where a defendant who confessed will be found guilty, participants in Experiment 1 overwhelmingly believed the suspect to be guilty, regardless of how inconsistently the confession was given. However, in Experiment 2, we were able to show that participants who viewed inconsistent confessions that could be verified using a secondary source (the police report) had significantly reduced belief in guilt, with a nearly 50:50 split between guilty/not guilty verdicts. From this we can infer that although inconsistencies might make people concerned about the veracity of the confession, those concerns do not necessarily translate to a strong belief that the suspect is innocent. Therefore, while inconsistent eyewitness testimony might cause jurors to discount that witnesses' evidence as unreliable, inconsistencies in confession evidence appear to create a similar effect on the discounting of evidence, albeit a smaller one.

In line with previous studies (e.g., Jones et al., 2019; Palmer et al., 2016), the findings in Experiment 2 run counter to wrongful conviction cases in which juries erroneously accepted a false confession as proof of guilt, and indicate that individual jurors are capable of rejecting inconsistent confession evidence. Aside from the complexity of the body of evidence considered by jurors in court, the way the confession itself is presented may play a role in shaping the way evidence is processed by jurors. In jury studies, single-page confession statements are often used, which preclude the juror from seeing any negotiation between the suspect and interviewing officers over factual details that might trigger suspicion about the confession's veracity. Multi-page transcripts, such as the one used in our studies, may further allow participants to see where inconsistencies arose and formulate hypotheses as to why the confession was inconsistent. However, the argument for using one type of written stimulus over the other is not strong, as similar results have emerged across recent studies

using transcripts and confession statements, suggesting that the effects of inconsistencies on judgments of guilt are robust to such variation in evidence materials.

Video-recorded confessions have been used to assess accuracy at detecting true versus false confessions (e.g. Kassin, Meissner, & Norwick, 2005), and the increasing use of video-recording of interrogations raises the possibility that jurors might be asked to evaluate increasingly detailed confession evidence (e.g., by considering a recording or transcript of the entire interrogation procedure). Although the present research used materials that were more detailed than a confession statement, other materials that more closely mimic evidence presented in some courtrooms, such as video-recordings or cross examination transcripts, would increase ecological validity in this type of study. Similarly, with regards to a realistic level of trial complexity, our study involved individuals making judgments based on a single police interview, rather than multiple pieces of competing evidence (e.g., opening statements; cross-examination) over a period of days. Future research might use the approach utilised in other studies (Fein et al., 1997; Fein et al., 2010) where suspicion-raising and disambiguating evidence are sent separately to participants over a period of time to create a more complex and realistic set of stimulus materials.

The present study, along with previous research (Alceste et al., 2019; Henderson & Levett, 2016; Palmer et al., 2016; Woestehoff & Meissner, 2016), shows that individual jurors are capable of both noticing inconsistent confession information and, in some circumstances, reducing their belief in a suspect's guilt due to decreased confidence in the veracity of the confession. However, juries function as a collective, rather than as individuals, and future research should consider the way that dissenting jurors make verdict decisions based on inconsistent confession evidence. As seen in the case of Pedro Hernandez, at one stage during the deliberations the verdict could have swung either way, although the trial ended with one holdout juror refusing to move from a not guilty verdict.

Future research into the effects of deliberation would contribute to understanding why individual jurors are sometimes sceptical of inconsistent confessions, and yet juries are still convicting based on such evidence.

Finally, although this research helps us understand when and why jurors are sometimes sceptical about confessions that contain inconsistencies, it also raises a new question: How do we best use this knowledge to help jurors process confession evidence? Our research shows that generating a plausible alternative explanation for a confession can lead to reduced judgments of guilt. Such scepticism about the validity of an inconsistent confession might help avoid wrongful convictions in cases involving questionable confession evidence (e.g., confessions that contain errors or were obtained via coercion), but might impede justice in cases involving unproblematic confession evidence (e.g., a voluntary confession obtained under best investigative practices and accompanied by convincing corroborating evidence). Further research is needed to determine how to best equip jurors to be discerning in their evaluation of confession evidence.

Summary and Conclusions

These studies build on existing research into how jurors might process confession evidence, by systematically investigating contradictory findings of the effects of inconsistencies in confession evidence on belief in guilt (Malloy & Lamb, 2010; Palmer et al., 2016). Specifically, our studies investigated the novel area of the effects of plausibility of self-generated alternate explanations for an inconsistent confession and found that plausibility played an important role in the relationship between confession inconsistencies and judgments of suspect guilt. In line with recent research (Alceste et al., 2019; Henderson & Levett, 2016; Palmer et al., 2016; Woestehoff & Meissner, 2016), the findings of the present studies paint an optimistic view that jurors can be discerning when evaluating confession evidence, and extend our understanding of the mechanisms behind decisions about suspect

guilt when that confession evidence is inconsistent. Our combined research concludes that although people are capable of scrutinizing confession evidence more closely than previously thought, not all types of inconsistencies will reduce perception of the suspect's guilt. Similar to Woestehoff and Meissner (2016), participants in our studies showed both the capacity to identify problematic confessions, and to imagine why an innocent person might confess in that situation. However, to sway belief in guilt, solid evidence was required in the form of factual errors, with contradictions proving too inconsequential to be of influence. Factual errors play a crucial role, because they prompt the consideration of alternative explanations for the confession which, in turn, can reduce perceptions of likely guilt. The promising notion of discerning jurors encourages continued investigation into the elements of confession evidence that might dissuade jurors from automatically accepting confessions and making a more considered assessment of their evidentiary value.

References

Alceste, F., Crozier, W. E., Strange, D. (2019). Contaminated confessions: How source and consistency of confession details influence memory and attributions. *Journal of Applied Research in Memory and Cognition*, *8*, 78-91. doi:10.1016/j.jarmac.2018.08.005

Appleby, S. C., & Kassin, S. M. (2016). When self-report trumps science: Effects of confessions, DNA and prosecutorial theories on perceptions of guilt. *Psychology, Public Policy, and Law, 22*(2), 127-140. doi:10.1037/law0000080

Bang, B. L., Stanton, D., Hemmens, C., & Stohr, M. K. (2018). Police recording of custodial interrogations: A state-by-state legal inquiry. *International Journal of Police Science and Management, 20*(1), 3-18. doi:10.1177/1461355717750172

Berman, G. L., Narby, D. J., & Cutler, B. L. (1995). Effects of inconsistent eyewitness statements on mock-jurors' evaluations of the eyewitness, perceptions of defendant culpability and verdicts. *Law and Human Behavior, 19*(1), 79-88. doi:10.1007/BF01499074

Bornstein, B. H. (1999). The ecological validity of jury simulations: Is the jury still out? Law and *Human Behavior, 23,* 75–91. doi:10.1023/A:1022326807441

Brewer, N., & Hupfeld, R. M. (2004). Effects of testimonial inconsistencies and witness group identity on mock-juror judgments. *Journal of Applied Social Psychology, 34*(3), 493-513. doi:10.1111/j.1559-1816.2004.tb02558.x

Fein, S., McCloskey, A. L., & Tomlinson, T. M. (1997). Can the Jury Disregard that Information? The Use of Suspicion to Reduce the Prejudicial Effects of Pretrial Publicity and Inadmissible Testimony. *Personality and Social Psychology Bulletin, 23*(11), 1215-1226. doi:10.1177/01461672972311008

Fein, S., Morgan, S. J., Norton, M. I., & Sommers, S. R. (2010). Hype and suspicion: The effects of pretrial publicity, race, and suspicion on jurors' verdicts. *Journal of Social Issues, 53*(3), 487-502. doi:10.1111/j.1540-4560.1997.tb02124.x

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin, 117*, 21-38. doi:10.1037/0033-2909.117.1.21

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York: The Guildford Press.

Henderson, K. S., & Levett, L. M. (2016). Can expert testimony sensitize jurors to variations in confession evidence? *Law and Human Behavior, 40*(6), 638-649. doi:10.1037/lhb0000204

Henkel, L. A., Coffman, K. A. J., & Dailey, B. A. E. M. (2008). A survey of people's attitudes and beliefs about false confessions. *Behavioral Sciences & the Law, 26*(5), 555-584. doi:10.1002/bsl.826

Innocence Project. (2017). Retrieved from http://www.innocenceproject.org/causes/false-confessions-admissions/

Jones, E. E. [Edward], & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. *Advances in Experimental Social Psychology, 2*, 219-266. doi:10.1016/S0065-2601(08)60107-0

Jones, E. E. [Eric], Bandy, A. D., & Palmer Jr., P. G. (2019) "I did it, but not like that": Effects of factually incorrect confessions on juror judgments. *Psychiatry, Psychology and Law*, *26*, 553-570. doi:10.1080/13218719.2018.1519467

Kassin, S. M. (2008). False confessions: Causes, consequences, and implications for reform. *Current Directions in Psychological Science, 17*(4), 249-253. doi:10.1111/j.1467-8721.2008.00584.x

Kassin, S. M. (2012). Why confessions trump innocence. *American Psychologist, 67*(6), 431-445. doi:10.1037/a0028212

Kassin, S. M., Drizin, S. A., Grisso, T., Gudjonsson, G. H., Leo, R. A., & Redlich, A. D. (2010). Police-induced confessions, risk factors and recommendations: Looking ahead. *Law and Human Behavior, 34*(1), 49-52. doi:10.1007/s10979-010-9217-5

Kassin, S. M., Meissner, C. A., & Norwick, R. J. (2005). "I'd know a false confession if I saw one: A comparative study of college students and police investigators". *Law and Human Behavior*, *29*(2), 211-227. doi: 10.1007/s10979-005-2416-9

Kassin, S. M., & Neumann, K. (1997). On the power of confession evidence: An experimental test of the fundamental difference hypothesis. *Law and Human Behavior, 21*(5), 469-484. doi:10.1023/A:1024871622490

Kassin, S. M., & Wrightsman, L. S. (1981). Coerced confessions, judicial instruction, and mock juror verdicts. *Journal of Applied Social Psychology, 11*(6), 489-506. doi:10.1111/j.1559-1816.1981.tb00838.x

Kelley, H. H. (1973). The process of causal attribution. *American Psychologist*, 107-128.

Kerr, N. L., & Bray, R. M. (2005). Simulation, realism and the study of the jury. In N. Brewer & K. D. Williams (Eds.), *Psychology and law: An empirical perspective* (pp. 322–364). New York: Guilford.

Leo, R. A. (2008). *Police Interrogation and American Justice*. New York: Harvard University Press.

Leo, R. A. (2013). Why interrogation contamination occurs. *Ohio State Journal of Criminal Law, 11*(1), 193-215.

Leo, R. A., & Ofshe, R. J. (1998). The consequences of false confessions: Deprivations of liberty and miscarriages of justice in the age of psychological interrogation. *The Journal of Criminal Law and Criminology, 88*(2), 429-496. doi:10.2307/1144288

Malloy, C., & Lamb, M. E. (2010). Biases in judging victims and suspects whose statements are inconsistent. *Law and Human Behavior, 34*, 46-48. doi:10.1007/s10979-009-9211-y

Neuschatz, J. S., Lawson, D. S., Swanner, J. K., Meissner, C. A., & Neuschatz, J. S. (2008). The effects of accomplice witness and jailhouse informants on jury decision making. *Law and Human Behavior, 32*(2), 137-149. doi:10.1007/s10979-007-9100-1

Palmer, M. A., Button, L., Barnett, E., & Brewer, N. (2016). Inconsistencies undermine the credibility of confession evidence. *Legal and Criminological Psychology, 21*(1), 161-173. doi:10.1111/lcrp.12048

Palmer, M. A., Horry, R., & Brewer, N. (2011). Court relevant research methods. In K. Sheldon, J. Davies, & K. Howells (Eds.), *Research in practice for forensic professionals*. Devon, UK: Willan Publishing.

Redlich, A. D., Ghetti, S., & Quas, J. A. (2008). Perceptions of children during a police interview: A comparison of alleged victims and suspects. *Journal of Applied Psychology, 38*(3), 705-735. doi:10.1111/j.1559-1816.2007.00323.x

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. *Advances in Experimental Social Psychology, 10*, 173-220. doi:10.1016/S0065-2601(08)60357-3

Tenney, E. R., MacCoun, R. J., Spellman, B. A., & Hastie, R. (2007). Calibration trumps confidence as a basis for witness credibility. *Psychological Science, 18*(1), 46-50. doi:10.111/j.1467-9280.2007.01874.x

Woestehoff, S. A., & Meissner, C. A. (2016). Juror sensitivity to false confession risk factors: Dispositional vs. situational attributions for a confession. *Law and Human Behavior, 40*, 564-579. doi:10.1037/lhb0000201

Table 1

Means, standard deviations, and 95% confidence intervals for Experiment 1 measures

	n	M *(SD)*	95% CI
Belief in Guilt			
Consistent	31	5.83 *(3.75)*	4.45, 7.21
Contradictions	42	5.49 *(4.23)*	4.17, 6.81
Perception of Consistency *			
Consistent	29	7.94 *(1.50)*	7.37, 8.51
Contradictions	39	3.71 *(2.10)*	3.03, 4.38
Perception of Voluntariness			
Consistent	31	7.06 *(1.97)*	6.33, 7.78
Contradictions	42	6.92 *(1.98)*	6.30, 7.53
No. of alternate explanations			
Consistent	31	1.71 *(1.66)*	1.10, 2.32
Contradictions	42	1.45 *(1.48)*	.99, 1.92
Plausibility of favored alternate explanation			
Consistent	30	43.27 *(29.94)*	32.09, 54.45
Contradictions	42	39.86 *(26.57)*	31.58, 48.14

Note. * indicates significant difference between group means

Table 2

Means and standard deviations for Experiment 2 measures

	n	M *(SD)*	95% CI
Belief in Guilt *			
Consistent	44	5.58 *(4.77)*	4.23, 7.03
Factual Errors	45	.27 *(7.10)*	-1.87, 2.40
Perception of Consistency *			
Consistent	44	7.72 *(1.54)*	7.25, 8.18
Factual Errors	45	3.47 *(2.39)*	2.75, 4.19
Perception of Voluntariness			
Consistent	44	7.46 *(2.30)*	6.76, 8.15
Factual Errors	45	6.80 *(2.45)*	6.06, 7.54
No. of alternate explanations			
Consistent	44	1.32 *(1.65)*	.82, 1.82
Factual Errors	45	1.58 *(1.41)*	1.16, 2.00
Plausibility of favored alternate explanation*			
Consistent	44	30.36 *(34.16)*	19.98, 40.75
Factual Errors	45	45.76 *(29.60)*	36.86, 54.65

Note. 95% CI = 95% confidence intervals, * = significant difference between group means

Figure Captions

Figure 1. The effect of confession inconsistency on belief in guilt, mediated by belief in a self-generated alternate explanation for the confession (other than guilt) in Experiment 2. * ≤.05, **≤.01, ***≤.001

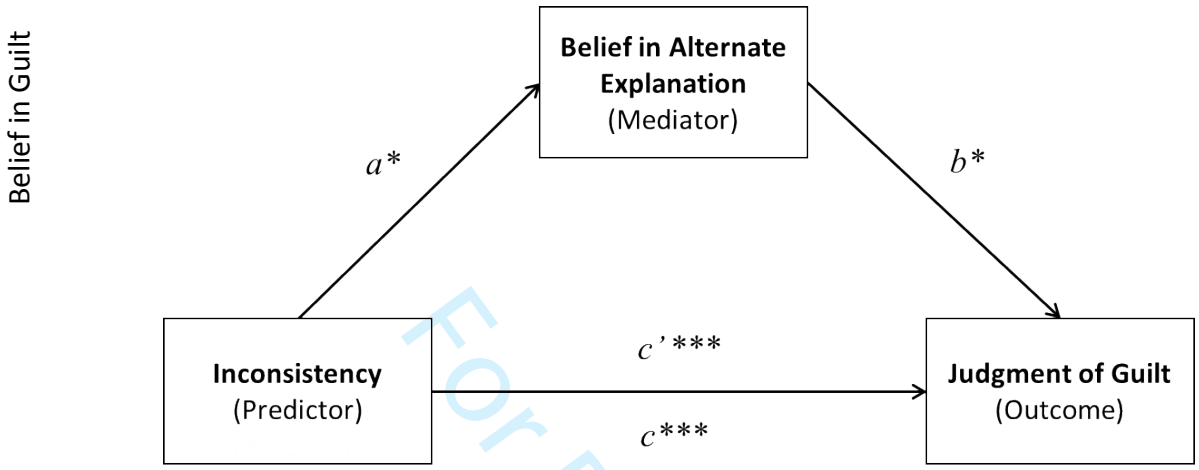

Figure 1. The effect of confession inconsistency on belief in guilt, mediated by belief in a self-generated alternate explanation for the confession (other than guilt) in Experiment 2. * ≤.05, **≤.01, ***≤.001